

DOCUMENT AUTOMATIC CLASSIFICATION SYSTEM, UNNECESSARY WORD DETERMINATION METHOD AND DOCUMENT AUTOMATIC CLASSIFICATION METHOD

5

ABSTRACT

It is an object of the present invention to eliminate unnecessary words effectively in document automatic classification.

10 A document automatic classification system comprising a classified document set storage device 21 for storing documents classified according to category, a category table generation unit 31 for generating a table broken down by category including information on a frequency of appearance of a word contained in a document acquired from the classified document set storage device 21, an unnecessary word determination and elimination unit 32 for eliminating an unnecessary word for each category from the table on the basis of a frequency of appearance in each category of a given word acquired from the table broken down by category generated by the category table generation unit 31, a classification catalog storage device 22 for storing the table from which the unnecessary word was eliminated by the unnecessary word determination and elimination unit 32, a classification target document storage device 23 for storing documents to be classified, and a document classification processing unit 33 for classifying the documents to be classified stored in the classification target document storage device 23 by using the table stored in the classification catalog storage device 22.

15

20

(Selected Drawings) Fig. 1

25